

## 2021 National Nursing Research Roundtable

# Mining EHR and Population Level Data Sources to Assess the Impact of Social and Behavioral Determinants of Health on Healthcare Utilization and Health Outcomes

**Elham Hatef, MD, MPH, FACPM**

*Center for Population Health IT*

*General Preventive Medicine Residency Program*

*Department of Health Policy and Management*

*March 5<sup>th</sup>, 2021*



**JOHNS HOPKINS**  
BLOOMBERG SCHOOL  
of PUBLIC HEALTH

# Introduction

- Development of **EHRs & Cross-Provider Regional HIE** (CRISP)
  - Opportunity for **Data-Oriented SDOH Assessment & Intervention**
    - The **point of care** (assessment & referrals of an individual with social need)
    - The **health delivery system level** (hiring a social worker in the clinic)
    - The **community** (building or strengthening community-based initiatives)

Population Health Management, Ahead of Print |

 Full Access

# Integrating Social and Behavioral Determinants of Health into Population Health Analytics: A Conceptual Framework and Suggested Road Map

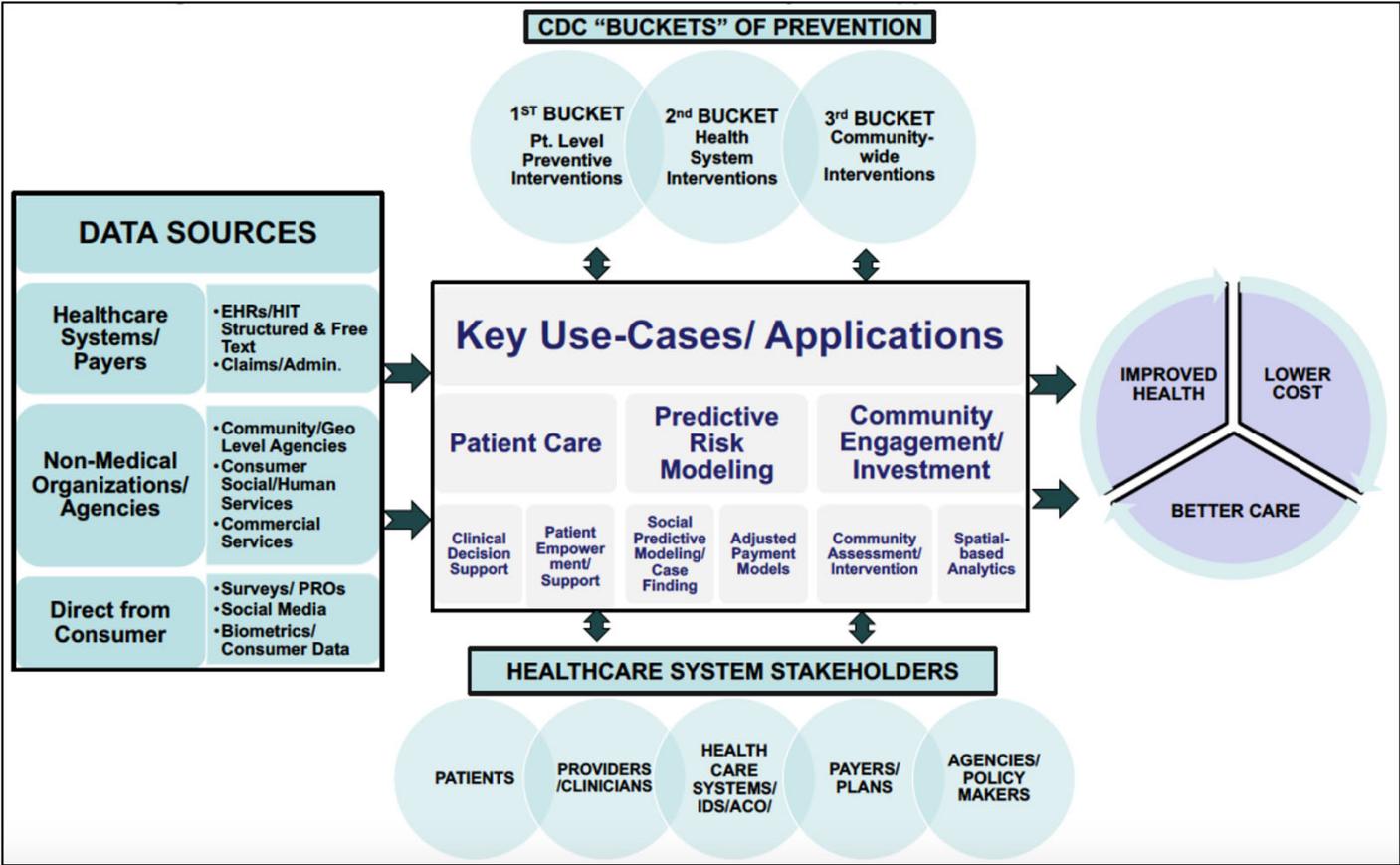
Zachary Predmore, Elham Hatef, and Jonathan P. Weiner 

Published Online: 13 Mar 2019 | <https://doi-org.proxy1.library.jhu.edu/10.1089/pop.2018.0151>

 Sections  View Article

 Tools  Share

# SDOH-Integrated Framework



# Data Sources: Healthcare Systems

- **Electronic Health Records**
  - Structured
  - Unstructured (Free-Text)

Currently submitted to: [JMIR Medical Informatics](#)

Date Submitted: Feb 22, 2019

Open Peer Review Period: Feb 22, 2019 - Mar 1, 2019

(closed for review but you can still tweet)

[Tweet](#)

NOTE: This is an **unreviewed** Preprint

Preprint

Assessing the Availability of Social and Behavioral Determinants Data in Structured and Unstructured Electronic Health Records: A Retrospective Analysis of a Multi-Level Healthcare System

Elham Hatef; Masoud Rouhizadeh; Iddrisu Tia; Elyse Lasser; Felicia Hill-Briggs; Jill Marsteller; Hadi Kharrazi

# Data Sources: Healthcare Systems

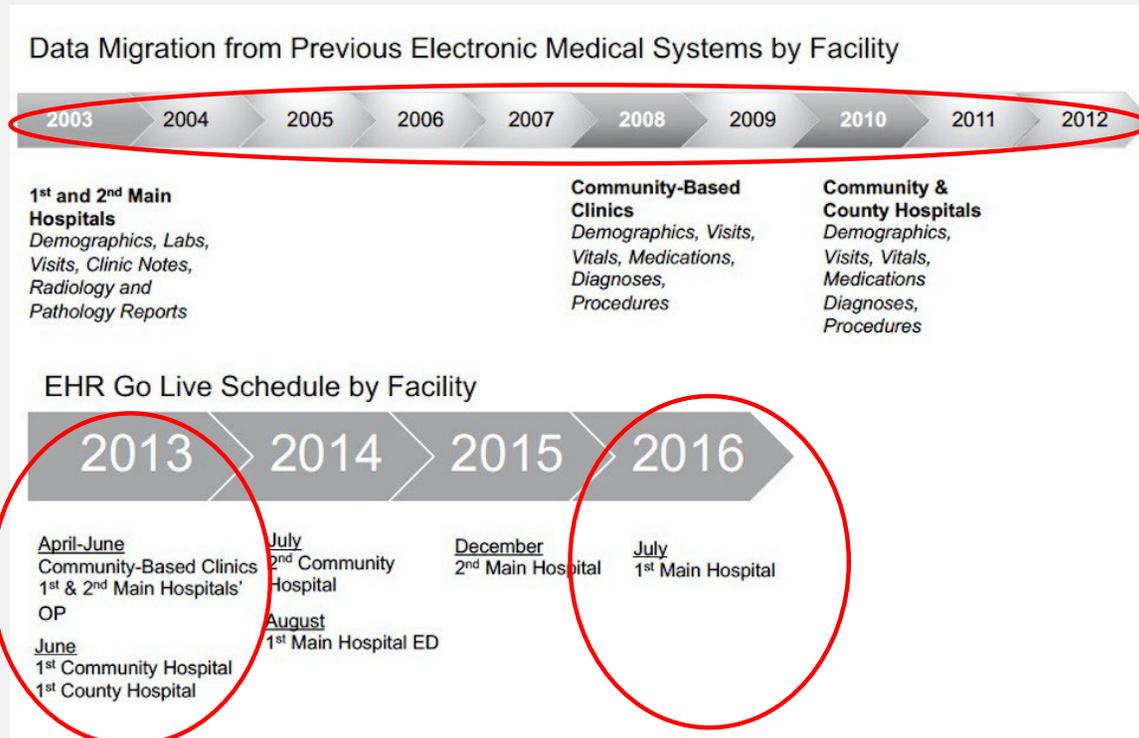
- We used Johns Hopkins EHR data from 2003 to 2018

- **Structured Data**

January 2003 and June 2018 from 5,401,324 unique patients

- **Unstructured Data**

July 2016 and May 2018 of 1,188,202 unique patients



# Collection Methods & Characteristics of Selected SBDH in Structured Data

Common Collection Method	Completeness Rate	Collection Date	Facility Type	History and Details	Other Collection Methods <sup>b</sup>
<b>Patient Address / Zip Code</b>					
Upon registration of each encounter. Documented as a street name & number, an optional line for apartment or other information, a city, a state or province, and a zip code.	<b>~5.2 million patients (95%)</b>	2003-Current	All facilities at the time of registration	~66% of patients' address change records are available, with effective start and end dates to track address change over time	Billing Address, Claims Processing Address, Home Health Encounters and Episodes, Communications for Specific Encounters
<b>Ethnicity</b>					
Upon registration of each encounter	<b>~2.7 million patients (50%)</b>	2003-Current	All facilities at the time of registration	Ethnicity (Hispanic or Not Hispanic) captured separately from race	Transplant Organ Donors, Ethnicity Questionnaire, Ethnicity Origin Questionnaire
<b>Race</b>					
Upon registration of each encounter	<b>~4.9 million patients (90%) indicated at least one race</b>	2003-Current	All facilities at the time of registration	Patients can self-identify multiple races	Home Health, Transplant Organ Donors
<b>Preferred Language</b>					
At the time of admission	<b>2,718,416 patients (50%)</b>	2003-Current	All facilities at the time of an encounter	The top preferred languages, by unique patient count: English (2,626,379, 48.6%) & Spanish (53,446, 0.9%) <sup>c</sup>	Flowsheets, Questionnaires, Clinical Notes
<b>Alcohol Use: "Alcoholic Drinks Per Week"</b>					
Social History portion of EHR during a patient encounter, whether in-person or not in-person encounters (telephone, MyChart <sup>d</sup> , documentation)	<b>490,348 (9.08%) patients - 178,789 (3.31%) patients reported one or more drinks per week</b>	2013-Current	All facilities at the time of an encounter	Reports show having any value (including 0 alcoholic drinks per week) in social history	Flowsheets, Questionnaires, Clinical Notes
<b>Smoking Status</b>					
Social History portion of EHR during a patient encounter, whether in-person or not in-person encounters (telephone, MyChart <sup>d</sup> , documentation)	<b>1,728,749 (32%) patients reported having any value smoking status in social history</b>	2013-Current	All facilities at the time of an encounter	Smoking Quit Date is also populated, but only in 137,958 (2.6%) of encounters <sup>e</sup>	Flowsheets, Questionnaires, Clinical Notes

## Number of Patients with Selected SDOH Domains in EHR – Using Diagnoses-Based Query

SBDH Categories and Subtypes/Codes	Diagnoses-Based Query
<b>Social Connection / Isolation</b>	<b>35,171 (0.64%)</b>
Z60.2 Problems Related to Living Alone	1222
Z60.4 Social Exclusion and Rejection	223
Z63.0 Relationship Problems (with spouse/ partner)	852
Z63.5 Family Disruption (separation/ divorce)	548
Z63.8 Other Primary Support Group Problems	2230
Z63.9 Unspecified Primary Support Group Problem	3247
Z65.9 Unspecified Psychosocial Circumstances	938
Z73.4 Inadequate Social Skills	81
Z91.89 Other Specified Personal Risk Factors	18,947
R45.8 Other Emotional State Symptoms and Signs	3340
<b>Housing Issues</b>	<b>10,433 (0.19%)</b>
Z59.0 Homelessness	7022
Z59.1 Inadequate Housing	120
Z59.8 Other Housing Problems	3291
<b>Income / Financial Resource Strain</b>	<b>3543 (0.06%)</b>
Z59.5 Extreme Poverty	68
Z59.6 Low Income	72
Z59.7 Insufficient Social Insurance and Welfare	46
Z59.8 Other Economic Circumstances Problems	3357

## Study Updates - Characteristics of Study Population and Those with Social Needs

	Total	Any Social Needs	Financial Challenges	Food Insecurity	Housing Issues	Transportation Issues
<b>Gender</b>						
Female	767,901 (58.3%)	17,687 (2.3%)	2,384 (0.3%)	830 (0.1%)	3,312 (0.4%)	15,990 (2.1%)
Male	549,434 (41.7%)	11,250 (2%)	804 (0.1%)	430 (0.1%)	2,002 (0.4%)	9,654 (1.8%)
<b>Age</b>						
18-34	299,220 (22.7%)	8,440 (2.8%)	1,281 (0.4%)	284 (0.1%)	1,895 (0.6%)	7,587 (2.5%)
35-44	205,697 (15.6%)	6,040 (2.9%)	582 (0.3%)	201 (0.1%)	1,010 (0.5%)	5,429 (2.6%)
45-54	199,700 (15.2%)	4,787 (2.4%)	315 (0.2%)	185 (0.1%)	688 (0.3%)	4,249 (2.1%)
55-64	232,253 (17.6%)	4,508 (1.9%)	476 (0.2%)	218 (0.1%)	947 (0.4%)	3,831 (1.6%)
65-74	195,356 (14.8%)	3,181 (1.6%)	343 (0.2%)	212 (0.1%)	541 (0.3%)	2,777 (1.4%)
75-84	115,758 (8.8%)	1,478 (1.3%)	156 (0.1%)	99 (0.1%)	187 (0.2%)	1,341 (1.2%)
85+	69,351 (5.3%)	503 (0.7%)	35 (0.1%)	61 (0.1%)	46 (0.1%)	430 (0.6%)
<b>Race</b>						
Caucasian	741,472 (56.3%)	14,829 (2%)	599 (0.1%)	585 (0.1%)	1,523 (0.2%)	13,359 (1.8%)
African American	341,439 (25.9%)	8,168 (2.4%)	967 (0.3%)	446 (0.1%)	2,045 (0.6%)	6,684 (2%)
Asian American	71,787 (5.4%)	2,098 (2.9%)	27 (0%)	118 (0.2%)	52 (0.1%)	1,959 (2.7%)
Pacific Islander	1,338 (0.1%)	44 (3.3%)	3 (0.2%)	1 (0.1%)	5 (0.4%)	41 (3.1%)
Other	103,903 (7.9%)	3,230 (3.1%)	1,579 (1.5%)	89 (0.1%)	1,653 (1.6%)	3,073 (3%)
<b>Location</b>						
Baltimore County	252,957 (19.2%)	4,111 (1.6%)	962 (0.4%)	221 (0.1%)	1,402 (0.6%)	3,475 (1.4%)
Howard	152,471 (11.6%)	3,648 (2.4%)	41 (0%)	53 (0%)	141 (0.1%)	3,505 (2.3%)
Montgomery	259,383 (19.7%)	6,298 (2.4%)	54 (0%)	437 (0.2%)	202 (0.1%)	5,730 (2.2%)
Prince George's	82,953 (6.3%)	1,932 (2.3%)	41 (0%)	30 (0%)	104 (0.1%)	1,844 (2.2%)
Baltimore city	230,214 (17.5%)	5,858 (2.5%)	1,897 (0.8%)	359 (0.2%)	2,991 (1.3%)	4,435 (1.9%)

## Study Updates - Breakdown of SDOH Domains by Healthcare Utilization

	Outpatient	Emergency Department	Inpatient
<b>Total Study Population</b>	m=3.458 (sd=8.73)	m=0.159 (sd=0.636)	m=0.061 (sd=0.241)
<b>Any Social Needs</b>	m=8.277 (sd=15.424)	m=0.495 (sd=3.038)	m=0.137 (sd=0.473)
<b>Financial Challenges</b>	m=16.299 (sd=25.126)	m=0.626 (sd=3.263)	m=0.313 (sd=0.572)
<b>Food Insecurity</b>	m=18.905 (sd=26.856)	m=0.881 (sd=4.309)	m=0.368 (sd=0.783)
<b>Housing Issues</b>	m=14.849 (sd=22.263)	m=1.963 (sd=6.76)	m=0.4 (sd=0.814)
<b>Transportation Issues</b>	m=7.436 (sd=14.029)	m=0.207 (sd=1.323)	m=0.096 (sd=0.356)
<b>Patients with the 3 most coded social needs</b>			
Z59.8 (Other problems related to housing and economic circumstances)	m=15.956 (sd=24.012)	m=0.62 (sd=3.293)	m=0.313 (sd=0.574)
Z59.0 (homelessness)	m=14.357 (sd=20.798)	m=5.129 (sd=11.267)	m=0.662 (sd=1.159)
Z76.89 (encountering health services in other circumstances)	m=6.326 (sd=11.645)	m=0.161 (sd=1.317)	m=0.068 (sd=0.306)

# Data Sources: Healthcare Systems

- **NLP Process to Identify SDOH in EHR Unstructured Data**
  - Used text mining techniques such as pattern matching
  - To craft the linguistic patterns, an expert team focused on three domains
    - Social Connection/Isolation
    - Housing Issues
    - Income / Financial Resource Strain

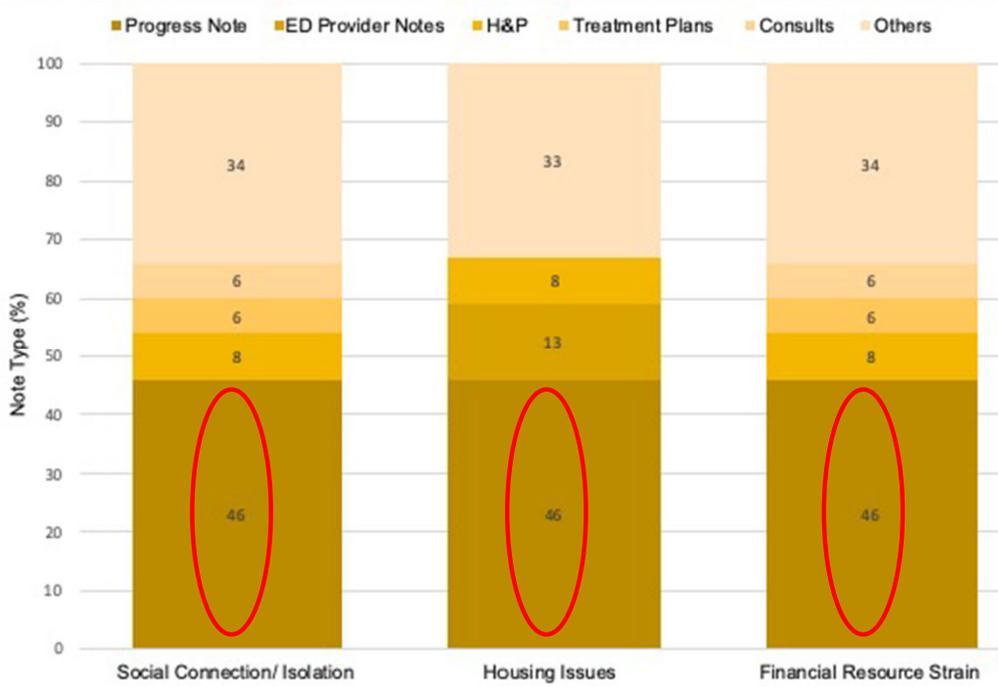
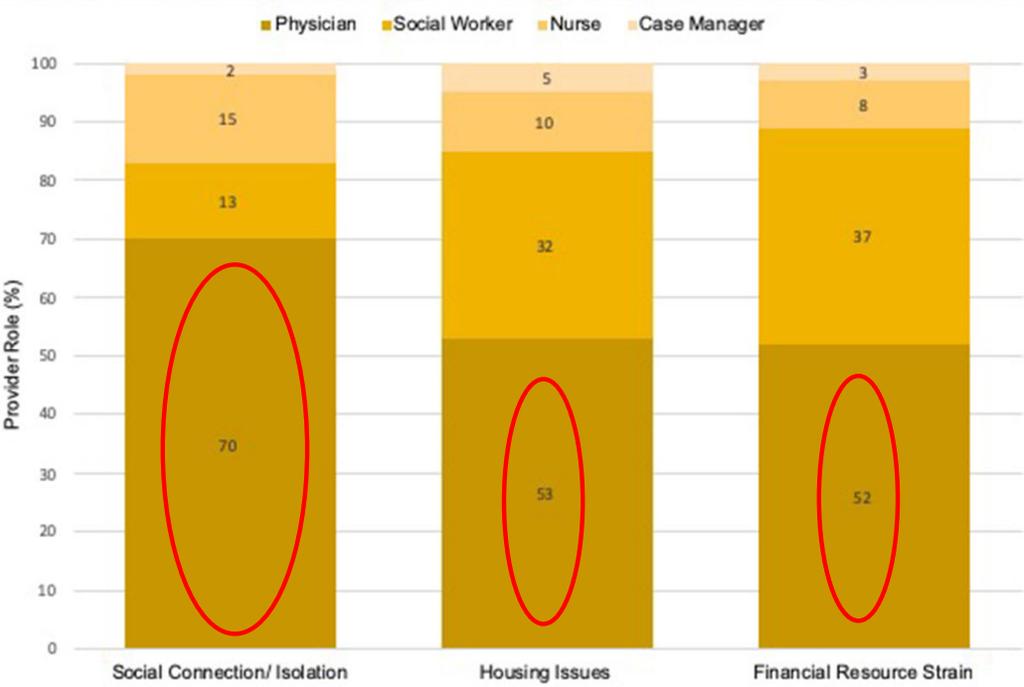
# Data Sources: Healthcare Systems

- **NLP Process to Identify SDOH in EHR Unstructured Data**
  - To develop linguistic patterns
    - Developed a comprehensive list of all available codes
      - ICD-10, CPT, LOINC codes, SNOMED terminologies
    - Reviewed description of SDOH in public health surveys & instruments
    - Reviewed phrases derived from a literature review of other studies
    - Used phrases identified through manual annotation of notes in a past study

# Number of Patients with Selected SBDH Domains in EHR – Using Diagnoses-Based Query and Unstructured Data

SBDH Categories and Subtypes/Codes	Diagnoses-Based Query	Unstructured
	Patient Count (%)	Patient Count (%)
<b>Social Connection / Isolation</b>	<b>35,171 (0.64%)</b>	<b>30,893 (2.59%)</b>
Z60.2 Problems Related to Living Alone	1222	-
Z60.4 Social Exclusion and Rejection	223	-
Z63.0 Relationship Problems (with spouse/ partner)	852	-
Z63.5 Family Disruption (separation/ divorce)	548	-
Z63.8 Other Primary Support Group Problems	2230	-
Z63.9 Unspecified Primary Support Group Problem	3247	-
Z65.9 Unspecified Psychosocial Circumstances	938	-
Z73.4 Inadequate Social Skills	81	-
Z91.89 Other Specified Personal Risk Factors	18,947	-
R45.8 Other Emotional State Symptoms and Signs	3340	-
<b>Housing Issues</b>	<b>10,433 (0.19%)</b>	<b>35,646 (2.99%)</b>
Z59.0 Homelessness	7022	-
Z59.1 Inadequate Housing	120	-
Z59.8 Other Housing Problems	3291	-
<b>Income / Financial Resource Strain</b>	<b>3543 (0.06%)</b>	<b>11,882 (0.99%)</b>
Z59.5 Extreme Poverty	68	-
Z59.6 Low Income	72	-
Z59.7 Insufficient Social Insurance and Welfare	46	-
Z59.8 Other Economic Circumstances Problems	3357	-

# Characteristics of EHR's Unstructured Data Containing SBDH – Stratified by Provider Role & Note Type



## **Collaborations with Other Health Systems - A Pilot Study to Improve the Use of Electronic Health Records for Identification of Patients with Social Needs: A Collaboration of Johns Hopkins Health System and Kaiser Permanente**

- Conducted independently, in a parallel and coordinated framework across sites
- The validation assessment and NLP algorithm logic were identical across sites
  - The “gold standard” for assessment of algorithm validity differed according to data availability
- Population Studied
  - Beneficiaries  $\geq 18$  years of age during 2016 through 2019 who received care at JHHS, KPMA, KPSCal

## Collaborations with Other Health Systems - A Pilot Study to Improve the Use of Electronic Health Records for Identification of Patients with Social Needs: A Collaboration of Johns Hopkins Health System and Kaiser Permanente

	JHHS	KPMAS	KPScaI
<b>Study Population (Patient No.)</b>	~1,200,000	~1,600,000	~4,700,000
<b>NLP Validation</b>			
Gold Standard Method	SDOH Questionnaire	SDOH Questionnaire	SDOH ICD codes Manual Annotation
<b>Sample Size</b>			
Patients/ Response No. (with/without residential Instability)	1,000 (500+/ 500-)	8,197 (833+,7364-)	300 (150+/150-)
Clinical Note No.	134,062	78,825	9,575
<b>NLP Algorithm Performance</b>			
Sensitivity	0.84	0.61	0.96
Specificity	0.96	0.87	0.97

# Use case: Predictive Risk Modeling

Johns Hopkins Health System

POPULATION HEALTH MANAGEMENT

Volume 00, Number 00, 2020

© Mary Ann Liebert, Inc.

DOI: 10.1089/pop.2020.0043

## Assessing the Impact of Social Needs and Social Determinants of Health on Health Care Utilization: Using Patient- and Community-Level Data

Elham Hatef, MD, MPH,<sup>1,2</sup> Xiaomeng Ma, MS,<sup>1</sup> Masoud Rouhizadeh, MS, PhD,<sup>3</sup>  
Gurmehar Singh, MS,<sup>1</sup> Jonathan P. Weiner, DrPH,<sup>1</sup> and Hadi Kharrazi, MD, PhD<sup>1,4</sup>

# Use case: Predictive Risk Modeling

- A **3-year retrospective population-based study**
  - Assessing individual- and community-level housing needs and how the level of housing needs impacts health care utilization
  - EHR data from July 2016 to May 2018
    - **1,187,956 unique patients** – EHR structured and unstructured data
  - 2017 U.S Census Data (American Community Survey)
    - Area Deprivation Index – A composite measure to rank neighborhoods based on their socio-economic challenges (higher ADI indicates more challenges)

# Key Findings

## Logistic Regression Assessing Factors Associated with Healthcare Utilization Among Johns Hopkins Patients Between 2016-2018

Variables	Overall Population			Medicaid Population		
	OR	95% CI	p-value	OR	95% CI	p-value
<b>Housing Issues</b>						
Homelessness	1.336	1.261, 1.416	<0.00001	1.902	1.576, 2.296	<0.00001
Housing Instability	1.489	1.380, 1.607	<0.00001	1.473	1.227, 1.769	<0.00001
Characteristics of the Building	0.888	0.818, 0.964	0.00469	0.847	0.640, 1.121	0.24500
<b>Age</b>						
	1.001	1.001, 1.002	<0.00001	1.010	1.008, 1.011	<0.00001
<b>Sex (male as reference)</b>						
Female	1.437	1.409, 1.467	<0.00001	1.563	1.458, 1.675	<0.00001
<b>Race (whites as reference)</b>						
African American	0.959	0.937, 0.981	0.00039	0.794	0.734, 0.858	<0.00001
<b>Neighborhood Characteristics (ADI National Rank, neighborhoods below the 10th percentile as reference)</b>						
Between 11 <sup>th</sup> & 89 <sup>th</sup> Percentiles	1.442	1.404, 1.481	<0.00001	1.466	1.239, 1.734	<0.00001
Above the 90 <sup>th</sup> Percentile	1.549	1.474, 1.627	<0.00001	1.598	1.325, 1.926	<0.00001
<b>Insurance Type (commercial insurance as reference)</b>						
Medicare	1.489	1.447, 1.532	<0.00001	-	-	-
Medicaid	2.078	1.997, 2.162	<0.00001	-	-	-
<b>Charlson Comorbidity Score (score of 0 as reference)</b>						
>= 3	55.444	53.333, 57.639	<0.00001	38.497	32.447, 45.675	<0.00001

# Challenges and Future Road Map

- **Lack of standards, tools and best practices**
- **Cost burden on providers and health systems**
- **Data interoperability, confidentiality, and validity**

# Conclusions

- **Growing pressures from payers and policy makers** to achieve greater value for patients and beneficiaries
- **Clinicians, health plans, and provider organizations** must, in the near term, find ways to more effectively introduce social and behavioral factors into the medical care process
- The need for **evidence & best practices** derived from the social, informatics, and public health sciences will be essential
- Although **numerous technical, operational, and political challenges remain**, there is little question that a social and behavioral determinant-enabled, **approach to patient care and population health** will be necessary

# Research Team & Collaborators

- **Center for Population Health IT**
  - Hsien-Yen Chang, PhD
  - Chris Kitchen, MA
  - Xiaomen Ma, MA
  - Kelly Searle, PhD
  - Elyse Lasser, MA
  - Hadi Kharrazi, MD, PhD
  - Jonathan Weiner, Dr.PH
- **Preventive Medicine Residency Program**
  - Marissa Tan, DO, MPH
  - Delaram Taghipoor, MD, MPH, MBA
- **Clinical Research Data Acquisition (CCDA)/ Institute for Clinical & Translational Research (ICTR)**
  - Masoud Rouhizadeh, PhD